DOCUMENT RESUME

ED 199 296 TM 810 309

AUTHOR Ory, John C.; Braskamp, Larry A.
TITLE Faculty Perceptions of the Quality and Usefulness of Three Types of Evaluative Information.
PUB DATE Aug 80
NOTE 15p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Faculty; *Course Evaluation; Data Collection; Faculty Promotion; Higher Education; Instructional Improvement; Simulation; *Student Evaluation of Teacher Performance; *Teacher Attitudes; Teacher Improvement; *User Satisfaction (Information)
IDENTIFIERS *Evaluation Utilization; *Instructor and Course Evaluation System

ABSTRACT

Faculty perceptions of simulated college student evaluative information collected by three methods were investigated: open-ended questions, group interviews, and objective questionnaire items. The objective items were from the Instructor and Course Evaluation System (ICES) student rating questionnaires. Open-ended questions were: (1) comment on the strengths and weaknesses of the instructor; (2) the beneficial aspects of the course; (3) suggestions for course improvement; and (4) comment on the grading procedures. The group interview process involved a pre- and post- interview session with the interviewer and the instructor. Faculty rated the three simulated evaluation reports on their potential for accuracy, trustworthiness, usefulness, comprehensiveness, believability, interpretability, and value as information used for self-improvement and promotion purposes. Faculty, in general, regarded the evaluative information to be more creditable, useful and accurate for their own self improvement than for promotion purposes. Faculty also desired more than one type of evaluative information regardless of the purpose of evaluation. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED199296

Faculty Perceptions of the Quality and Usefulness of Three Types of Evaluative Information

John C. Ory and Larry A. Braskamp
Measurement and Research Division
Office of Instructional Resources
University of Illinois
Urbana, IL 61801

August, 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. C. Ory
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

The purpose of the study was to investigate faculty perceptions of student evaluative information collected by three methods: objective questionnaire items, open-ended question, and group interviews. Faculty rated three simulated evaluation reports on their potential for accuracy, trustworthiness, usefulness, comprehensiveness, believability, interpretability, and value as information used for self-improvement and promotion purposes. Faculty, in general, regarded the evaluative information to be more credible, useful and accurate for their own self improvement than for promotion purposes. Faculty also desired more than one type of evaluative information regardless of the purpose of evaluation.

TM 810 301

Faculty Perceptions of the Quality and Usefulness of Three Types of Evaluative Information

Evaluation of the teaching competence of faculty is now commonplace in higher education (Centra, 1977). A number of different methods for collecting evaluative information about teaching effectiveness have been advocated (Brandenburg, Braskamp, Ory, 1980), including peer evaluations, self evaluations, and student group interviews as well as the most popular method of student rating scales. The research on the reliability and validity of student ratings has been voluminous (e.g., Feldman, 1977; Centra, 1980). Cohen (Note 1) in a meta-analysis of the relationship between student ratings and student achievement concluded that ratings of overall teaching were correlated with achievement (average r = .47), but were not as highly correlated with specific teachings such as rapport, interaction, feedback and evaluation. The convergent validity of differing types of evaluation information based on student group interviews, student objective rating items, written comments to open-ended questions has been demonstrated if the purpose is to assess overall teaching effectiveness (Ory, Braskamp, and Pieper, 1980; Braskamp, Ory, and Pieper, in press). There is also sufficient convergent validity among evaluations by peers, self, and students (e.g., Braskamp, Caulley, Costin, 1979) to warrant their use.

However, little is known about the credibility, trustworthiness, and usefulness of differing types of evaluative information from the perspective of the primary users, the faculty. The purpose of this study was to investigate faculty perceptions of student evaluative information collected by three methods--objective questionnaire items, written comments to open-ended questions, and group interviews. The major three questions asked in this study were:

1. Do faculty similarly rate the quality and usefulness of information if the purpose is for promotion or for self improvement?
2. Do faculty differentially rate the quality and usefulness of the three types of information when it is to be used for promotion?
3. Do faculty differentially rate the quality and usefulness of the three types of information when it is to be used for self improvement?

The evaluative information, based on three different methods of data collection, can be viewed as a message to a user. The conceptual framework, "Who says what to whom with what effects" used extensively in communication research was employed to interpret the results of this study. The "who" were students; the "how" was a written report, "to whom" were faculty receiving the evaluative information, and

the "effects" included various perceptions of the faculty about the accuracy, trustworthiness, comprehensiveness, believability, interpretability of the information and the usefulness of the information for their own self improvement as teachers and for their promotion. In this simulation experiment, the message on instructor competency varied by the format of the presentation of the data; i.e., the "what" differed by the type of information included in the report.

This framework has been used in a number of previous simulation experiments to investigate influences of source (who), message or content (what), and audiences (to whom). Educators in public shools regarded evaluation reports without data and jargon as more believable, practical, logical, and useful than reports with jargon and data (Thompson, Brown, Fergason, Note 3). Jargon-free but data-supported reports on testing were rated the least difficult by another group of educators (Brown, Braskamp, and Newman, 1978). Educators also expressed greater agreemnt with the evaluator's recommendations about the program evaluation if the report included graphs and percentages but not statistical inference statements by the evaluator (Brown and Newman, 1980).

Simulated Evaluation Reports

Three evaluation reports about an instructor and course were written from simulated data "collected" by three different methods: objective questionnaire items, open-ended questions, and group interviews. The objective items were from the Instructor and Course Evaluation System (ICES) student rating questionnaires (Office of Instructional Resources, Note 2). The ICES is a cafeteria-type student rating system that permits each instructor to select up to 23 items to be used on the questionnaire. The first 3 items on all student questionnaires are global items: "Rate the course content," "Rate the instructor," and "Rate the course in general." Students indicate their rating on these 3 items on a 6-point scale, with anchor points of "poor" and "excellent." The open-ended questions were the 4 standard questions printed on Side Two of the ICES questionnaire. The items are "Comment on the strengths and weaknesses of the instructor." "What are the beneficial aspects of the course?" "What suggestions do you have for course improvement?" and "Comment on the grading procedures."

Simulated responses to the objective questionnaire items were summarized in a computer-generated ICES "Instructor Report." Presented in the report are descriptive statistics of student responses to each of the three global and 23 instructor selected items. A comparison of the instructor's global

item ratings with the normative ratings of the instructor's department and of the university are also provided. The evaluation report developed from simulated written comments to open-ended questions included 4 hand-written samples of Side Two of the ICES questionnaire. The comments were written to present a general impression of instructor and course quality similar to the one depicted by the computerized Instructor Report. Both positive and negative comments about the course and instructor were included.

The third evaluation report included a description and results of a simulated student group interview conducted by staff members from the campus Office of Instructional Resources. The group interview process described in the simulation involved a pre- and post-interview session with the interviewer and the instructor. In the pre-interview session the instructor discussed teaching philosophy, course objectives and special concerns that can be discussed in the interview. Following the interview a 1-2 page summary report of student comments was written and discussed with the instructor in an informal session. In the simulation the faculty read a one page summary. Again, the general impression of instructor and course quality presented by the group interview report was similar to that communicated by the other two evaluation reports. Both strengths and weaknesses of the instructor and course were discussed.

Five judges independently read each of the three evaluation reports and rated the comparability of content presented in each report. Each of the judges, staff members of the campus Office of Instructional Resources, rated the content highly comparable across reports.

Data Collection

Three groups of 50 randomly selected university faculty members received one of the three simulated evaluation reports during the 1980 spring semester. Faculty were equally represented from each academic rank and discipline areas of agriculture, architecture, business, chemistry-biology, classics, education, fine arts, math, physical sciences and social sciences. The number of faculty returning the rating forms for each type of evaluation report were 25 (Objective Item Responses), 24 (Group Interview), and 22 (Written Comments).

Faculty were asked to rate the given report on seven dimensions: accuracy, trustworthiness, usefulness, interpretability, comprehensiveness, believability, and value as information. They rated the report on each dimension twice--first for its use for self improvement and second for promotion purposes. For each item the faculty indicated on a 5-point continuum from high (=1) to no amount

(=5) their rating on each dimension. Faculty were also asked to indicate if they would use the evaluation report for either purpose. A final question asked the faculty to recommend other types of evaluative information which they would like to receive in addition to the information presented in the given evaluation report.

Data Analysis

To analyze differences in faculty perceptions, a 3 x 2 multivariate analysis of variance (MANOVA) was computed on the seven scale items with type of evaluation report repeated across purpose. Post hoc univariate analyses of variance were computed to interpret significant MANOVA effects. Reported faculty use of the different evaluation reports was analyzed through a 1 x 3 ANOVA design using the F test statistic. An alpha of .05 was used as the level of statistical significance.

Results

The means and standard deviations for each of the first 7 scale items are presented in Table 1. Results of the MANOVA presented in Table 2 revealed significant ($p < .01$) differences in faculty ratings due to the purpose of the report and to the type of report for particular purposes. The univariate analyses of variance presented in Table 2 revealed that differences due to purpose of report were statistically significant for 6 of the 7 items. Faculty

Insert Tables 1 and 2

generally indicated that reports were potentially more accurate, trustworthy, useful, comprehensive, believable, and valuable when used for self-improvement rather than promotion purposes. Faculty also rated the group interview information as most comprehensive, followed in order by objective item responses and written student comments, regardless of the purpose of information.

However, each type of report did not receive the same relative ratings of trustworthiness and believability for both purposes. A statistically significant interaction of purpose by report type was obtained for items 2 and 6. As shown in figures 1 and 2, these interactions illustrate disordinate

Insert Figures 1 and 2

interactions, i.e., the rank order of the mean ratings of each type is not identical for both purposes. Of the three types of evaluative information, the ranking of the written student comments was first for self improvement but third for promotion on trustworthiness and believability.

Faculty use of the information in the evaluation report is summarized in Table 3. Reported faculty use of the information was not significantly different for type of evaluation report ($F\ [2,68] = 1.43$, $p < .30$). Approximately 67% of the faculty were willing to use the information regardless of type of evaluation report.

Insert Table 3

Faculty, when asked if they also wanted to receive the other types of evaluative information, generally desired more information. Of faculty receiving written comments information, 68.2% and 54.5% would have also appreciated receiving objective item responses and group interview summaries, respectively. Approximately eight percent of the faculty responding to the group interview report asked for objective item responses and written comments about their performance. Finally, 68% and 40% of the faculty rating the objective item response report also wanted to receive written comments and group interview results, respectively. Approximately one-third of all respondents also requested that some form of peer review be used to evaluate their performance.

Discussion

Based on this simulation experiment, the type of information presented--student written comments, computer printout of student response to fixed alternative questionnaire items, and written summary of student group interview--does influence the audience's (faculty) rating of the technical quality and credibility of the information. However the conceptual framework, "who says what how to whom with what effects" needs to be expanded to "under what conditions who says what how to whom with what effects." The conditions in this simulation refer to the purpose of the evaluation. In six of the seven ratings of the quality and credibility of evaluative information, faculty in general regarded the information to be more credible, useful, and accurate for their own self improvement than for promotion purposes.

However, faculty differentially rated the credibility (i.e., trustworthiness and believability) of the three types of information depending on the purpose. Faculty regarded student written comments as less credible than student responses to objective comments when the purpose was promotion but rated written comments as more credible when the purpose was self improvement. Perhaps the non-standardized, unique and personal written comments by students are perceived as too subjective for important personnel decisions. However, this highly idiosyncratic type of information about a particular course is viewed as useful diagnostic information for making course changes.

Finally, the faculty desired more than one type of evaluative information regardless of the purpose of evaluation. One third of the faculty also desired peer reviews, even though peer reviews are seldom done on this campus. Faculty in this simulation have supported the advice given by many in faculty evaluation (e.g., Doyle, 1975, Centra, 1980, Brandenburg, Braskamp, and Ory, 1980) that evaluation information about teaching competence should be collected from many different sources (e.g., peers, students) using a variety of differing methods of data collection.

Reference Notes

1. Cohen, P. A. A meta-analysis of the relationship between student ratings of instruction and student achievement. Unpublished doctoral dissertation. University of Michigan, Ann Arbor, Michigan, 1979.
2. Illinois Course Evaluation System: Its rationale and description (ICES Newsletter No. 2). Urbana-Champaign: University of Illinois, Measurement and Research Division, Office of Instructional Resources, August, 1977. (Mimeo)
3. Thompson, P., Brown, R. D., & Fergason, J. The impact of evaluation report styles on evaluation audiences. Paper presented at annual meeting of American Educational Research Association, San Francisco, April, 1979.

References

Brandenburg, D. C., Braskamp, L. A., & Ory, J. C. Considerations for an evaluation program of instructional quality, CEDR, 1980, 12, 8-12.

Braskamp, L. A., Caulley, D., & Costin, F. Student ratings and instructor self-ratings and their relationship to student achievement. American Educational Research Journal, 1979, 16, 295-306.

Braskamp, L. A., Ory, J. C., & Pieper, D. M. Student written comments: Dimensions of instructional quality. Journal of Educational Psychology, in press.

Brown, R. D., Braskamp, L. A., & Newman, D. L. Evaluator credibility and acceptance as a function of report styles: Do jargon and data make a difference? Evaluation Quarterly, 1978, 2, 331-341.

Brown, R. D. & Newman, D. L. Perceived need for evaluation as an influence on evaluation's impact on decision making. Educational Evaluation and Policy Analysis, in press.

Centra, J. A. Determining faculty effectiveness. San Francisco, CA: Jossey-Bass Publishers, 1980.

Centra, J. A. How universities evaluate faculty performances: A survey of department heads. (GREB No. 75-5 bc). Princeton, N.J.: Educational Testing Service, 1977.

Feldman, K. A. Consistency and variability among college students in rating their teachers and courses: A review and analysis. Research in Higher Education, 1977, 6, 223-274.

Ory, J. C., Braskamp, L. A., & Pieper, D. M. The congruency of student evaluative information collected by three methods. Journal of Educational Psychology, 1980, 72, 181-185.

Table 1
Scale Item Means and Standard Deviations

Type of Report	Self-Improvement Purposes X	Self-Improvement Purposes SD	Promotion Purposes X	Promotion Purposes SD
Item #1: How accurate?				
Group interview (n = 24)	2.56[1]	.73	2.94	1.06
Objective item responses (n = 25)	2.81	1.29	3.14	1.32
Written comments (n = 22)	2.53	.99	3.47	1.30
Item #2: How trustworthy?				
Group interview	2.63	.62	2.88	1.03
Objective item responses	3.05	1.12	3.19	1.17
Written comments	2.60	1.12	3.67	1.18
Item #3: How useful?				
Group interview	2.19	1.11	2.75	1.13
Objective item responses	2.86	1.46	3.10	1.26
Written comments	2.60	1.18	3.40	1.35
Item #4: How difficult to interpret?				
Group interview	3.19	1.05	3.25	1.07
Objective item responses	2.95	1.32	2.81	1.33
Written comments	3.60	1.12	2.93	1.67
Item #5: How comprehensive?				
Group interview	3.06	1.06	3.38	1.03
Objective item responses	3.72	1.10	3.81	1.17
Written comments	4.13	.83	4.33	.72
Item #6: How believable?				
Group interview	2.69	1.08	3.06	1.06
Objective item responses	2.95	1.16	3.10	1.09
Written comments	2.47	.99	3.33	1.35
Item #7: How valuable?				
Group interview	2.38	1.31	2.94	1.24
Objective item responses	2.91	1.45	3.14	1.39
Written comments	2.73	1.03	3.33	1.23

[1] Low scale values indicate high perceived accuracy, trustworthiness, usefulness, etc.

Table 2

MANOVA & ANOVA Summary Tables: Items 1-7

Source of Variation	SS	df	MS	F
MANOVA				
Type of Report (T)	35.26	14,497	2.52	1.12
Between Error	800.36		2.24	
Purpose of Report (P)	36.88	7,497	5.27	15.55*
Purpose x Type (PT)	12.70	14,497	.91	2.68*
Within Error	120.92		.34	
ANOVAS				
Item #1: How accurate?				
T	1.25	2	.62	.27
Between Error	114.48	68	2.34	
P	7.62	1	7.62	23.83**
PT	1.82	2	.91	2.84
Within Error	15.68	68	.32	
Item #2: How trustworthy?				
T	3.12	2	1.56	.80
Between Error	95.37	68	1.95	
P	6.02	1	6.03	19.36**
PT	4.16	2	2.08	6.68**
Within Error	15.25	68	.31	
Item #3: How useful?				
T	5.94	2	2.97	1.09
Between Error	133.95	68	2.73	
P	7.25	1	7.25	15.39**
PT	1.43	2	.71	1.51
Within Error	23.07	68	.47	
Item #4: How difficult to interpret?				
T	3.30	2	1.65	.58
Between Error	139.74	68	2.85	
P	1.58	1	1.58	3.79
PT	2.19	2	1.10	2.63
Within Error	20.42	68	.42	

Table 2, continued.

Source of Variation	SS	df	MS	F
Item #5: How comprehensive?				
T	16.01	2	8.00	4.24*
Between Error	92.45	68	1.39	
P	1.05	1	1.05	5.80*
PT	.22	2	.11	.60
Within Error	8.82	68	.18	
Item #6: How believable?				
T	.48	2	.24	.11
Between Error	108.18	68	2.21	
P	5.42	1	5.42	16.58**
PT	2.32	2	1.16	3.54*
Within Error	16.03	68	.33	
Item #7: How valuable?				
T	3.06	2	1.53	.50
Between Error	149.66	68	3.05	
P	5.55	1	5.55	17.35*
PT	.74	2	.37	1.16
Within Error	15.67	68	.32	

*p < .05

**p < .01

Table 3

Faculty Use and Preference for Student Use of Information

Item Responses		Objective Item Responses (n=25) f	%	Written Comments (n=22) f	%	Group Interview (n=24) f	%
Item #8: Would you use the information?							
Definitely would	(1)	7	28.0	7	31.8	10	41.7
	(2)	9	36.0	7	31.8	8	33.3
	(3)	4	16.0	4	18.2	5	20.8
	(4)	1	4.0	2	9.0	1	4.2
Definitely would not	(5)	4	16.0	2	9.0	0	0
		$\bar{X}$ = 2.44		$\bar{X}$ = 2.32		$\bar{X}$ = 1.88	
		SD = 1.39		SD = 1.29		SD = .90	
Item #9: Should students have access to the information to help them in course selection?							
Definitely should	(1)	8	32.0	4	18.2	9	37.5
	(2)	2	8.0	3	13.6	6	25.0
	(3)	5	20.0	6	27.3	4	16.7
	(4)	4	16.0	6	27.3	3	12.5
Definitely should not	(5)	6	24.0	3	13.6	2	8.3
		$\bar{X}$ = 2.92		$\bar{X}$ = 3.05		$\bar{X}$ = 2.29	
		SD = 1.61		SD = 1.33		SD = 1.33	

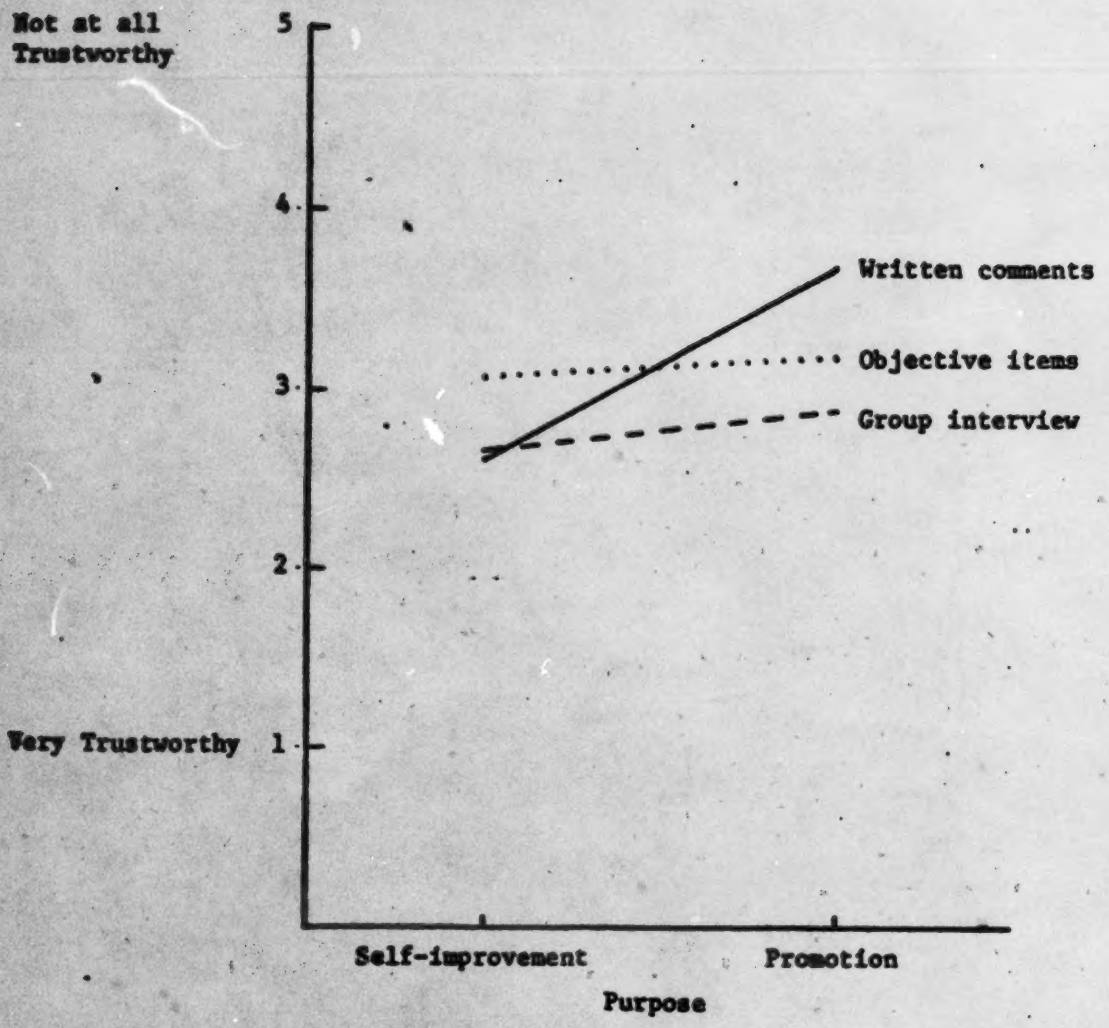

Figure 1. Purpose x type of report interaction for scale item #2: Trustworthiness

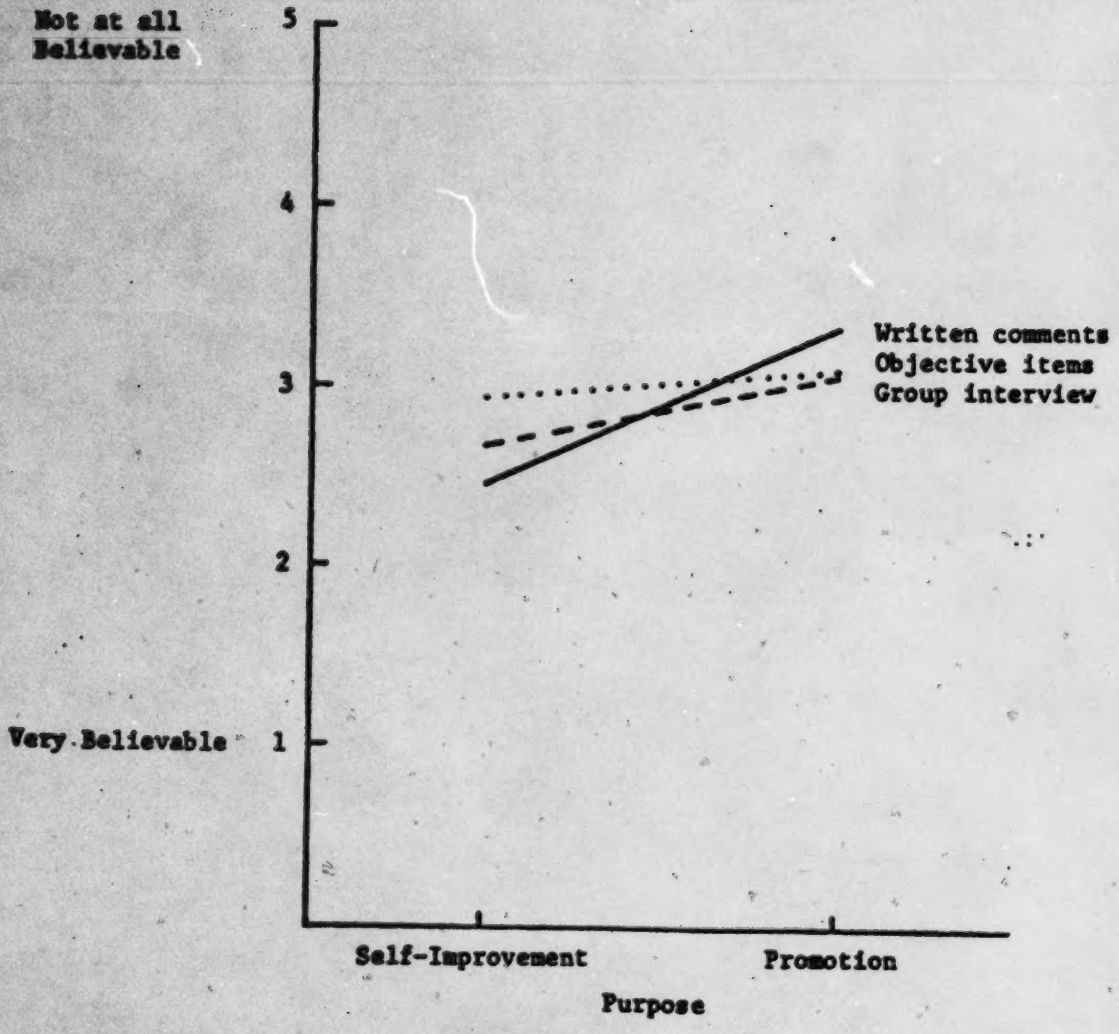

Figure 2. Purpose x type of report interaction for scale item #6: Believability